The Laughing Swing: Interacting with Non-Verbal Human Voice

Michal Rinott

School of Design
Holon Institute of Technology
Golomb 54, Holon, Israel
michalrin@gmail.com

ABSTRACT

This paper looks at non-speech uses of the human voice in interactive objects. A collection of projects using non-verbal voice, as input and as output, is briefly reviewed. The Laughing Swing - an interactive object using non-verbal voice as output, created by the author and associates - is described in terms of motivations, sound design, sonic behavior implementation and user responses. The significance and potential of interactions with non verbal voice is discussed.

1. INTRODUCTION

The human voice may well be the most significant stimulus in a human being's life. We hear it while still inside the womb, and create it when emerging into the world, after taking our first gasp of air. Long before speech is acquired, the voice is a premier tool of expression and interaction. The human infant uses vocal expressions of emotion: laughter, crying, exclamations of delight, distress, surprise; all of great communicative value, all immediately understandable to the listener, and all eliciting a strong emotional response.

Later in life, when speech is acquired and perfected, we do not replace these "raw" modes of expression with verbal references to our emotions, but rather intermix them – using language and paralanguage when we are more in control; weeping, laughing, gasping, moaning and shouting with joy when overcome with emotion.

1.1. Interactions with voice

The dream of interacting with computers using our natural voice (as portrayed, for example, in the 1960's Star Trek series and in Arthur C. Clark's Hal 9000), long preceded the technical ability to do so. Now coming of age, basic speech recognition (speech-to-text) and speech synthesis (text-to-speech) engines are so easily available that the are included with every new personal computer. Speaking commands to the computer, dictating text, and having responses spoken out in synthesized or recorded voice, are now employed in different contexts (cars, call centers and even home appliances). However, as successful or frustrating as this type of interaction may be, it covers only a small part of what the voice can produce, and the mind hear and understand.

1.2. Non-verbal interactions with voice

A number of recent projects in human-computer interaction have proposed interactions that employ modified speech and non-verbal uses of voice. In the voice-as-output direction, a notable example is "Spearcons" [1]. Spearcons are a form of auditory icon composed of a highly time-compressed presentation of the spoken phrase they represent. Very interestingly, Spearcons have been found to be more efficient than other forms of auditory icons, even when the resulting sound is no longer comprehensible as a particular word [1].

In a demonstration by Igarashi and Hughes [2], a combination of speech and non-verbal voice in used for input to a computer. In this interaction model, control of actions is performed by voice in a form called "voice-as-sound": the user changes a parameter by first naming it and then using continuous voice of a certain duration to set its value (for example, saying "Move up, Ahhhhhhhhhhh", will move the mouse pointer up as long as the voice continues). In related research within the context of assistive technologies, Sporka and his collaborators have demonstrated control over the mouse pointer and keyboard using humming and whistling [3, 4].

These interactions show initial potential for the use of non-verbal voice in interactions, in both the input and output directions. They begin to demonstrate that the speech can be deconstructed and manipulated, sped up and intermixed with non-verbal vocal utterances.

In the arts, attempts to deconstruct speech, to break the connection between vocal sounds and words, have an important place. To mention one, Luciano Berio's decomposition of Cathy Berberian's speech into sound in the piece "Ommagio a Joyce" (1958), and his later composition of moans, shouts, whimpers, cries, and laughter in "Visage" (1961), is considered one of the world's most influential works of electronic music [5]. It represents an important moment in the history of the sounding arts: a moment of enigmatic novelty and of new challenge for the listener.

In interactions with technological objects, values other than novelty and challenge are paramount. Efficiency and clarity, ease of use, objectivity and straightforwardness are usually sought in creating the human-computer dialog. In this context, it is obvious why verbal interactions, as input to the computer and as output from it, have taken center stage.

However, as our lives become infused with technology, and computers move from the office into our pockets and our everyday objects, the nature of our interactions with technology is changing. Their range is becoming broader, their purposes more specific, and their value is no longer measured only in terms of utility. There is room for more intriguing, enigmatic interactions, for less defined borders between functionality and engagement, utility and pleasure.

This creates an opportunity to further explore non-verbal interaction, and to move from efficient and dignified interactions to more engaging and emotional ones. So are we ready for our technological objects to moan, gasp, and laugh? Do we want to growl and shout at our machines? Perhaps the time is ripe. The next section describes in brief a number of such research projects.

1.3. Interactive objects using non-speech voice

"Blendie" is a voice controlled blender designed by Kelly Dobson at MIT Media Lab [6]. In order to control Blendie, the user must speak the language of the blender. Bobson describes the interaction such: "people induce the blender to spin by sounding the sounds of its motor in action. A person may growl low pitch blender-like sounds to get it to spin slow… and the person can growl blender-style at higher pitches to speed up". Dobson sees this project as representing a new form of connection with a machine, accessing and vitalizing the interplay of people and machines.
[BLENDIE.MOV]

"SoMo2 - the Speaking Mobile" is a mobile phone prototype designed by Graham Pullin and Crispin Jones at IDEO [7]. This phone allows its user to converse silently. Rather than speaking into the phone, a person receiving a call can respond with simple but expressive vowel sounds which are produced and intoned manually using a small joystick (e.g. yeeaaahhhh? [intonation: sarcastic], ooooohhhhhh... [intonation: hesitant], yeeaahh! [intonation: enthusiastic]). Pullin and Jones see this project as the "the antithesis of text messaging, in that it conveys rich emotional nuance at the expense of textual information".
[SOMO2.MOV]

"SonicTexing" is a handheld device that enables text entry using sound, created by Michal Rinott [8]. In SonicTexting letters are found and selected by navigating with a joystick and receiving synchronous auditory feedback: recorded sounds of letter phonemes (e.g. "Eeee" "Rrrrr"). The phonemes are sung in different pitches, loudness and frequency as a function of the user's navigation motions. Rinott sees this project as "an exploration of people's hand-ear coordination, and a proposal that through touch and sound, interacting with digital devices can become an experience on the borders between a functional task, an instrument and a game".
[SONICTEXTING.MOV]

"Audio Shaker" is a tactile container to capture, shake up and pour out sounds, created by Mark Hauenstein and Tom Jenkins [9]. Anything sung, spoken, clapped, whistled or played near this jug-like container is trapped inside. Caught sounds are transformed, given weight and permanence, reacting directly to the shaker's movements, subtle or violent. Hauenstein and Jenkins see this project as "an exploration of our perceptual understanding of sound; a rich, intuitive experience purposefully open to interpretation and imagination".
[AUDIO SHAKER.MOV]

1.4. The nature of objects that employ non-speech voice

Of the objects above, one employs voice as input (Blendie), two employ voice as output (SoMo2, SonicTexting) and one employs voice as both input and output (Audio Shaker). The growling-howling nature of the voice input in Blendie, and the physical malleability of sound by hands in the other three projects, contribute to an interaction that feels by nature less controlled, less predictable, less tame than the ones we are accustomed to with technology.

Except the Audio Shaker, the interactions described above enable a functional task to be performed (texting, "speaking", blending). Yet in their design, the creators stress interplay, emotion and nuance. It seems that a special quality is created when the object has a voice, and the user can manipulate it with hands or body. This voice, originally embodied in a person, then disembodied using recording technology or synthesis, is now re-embodied in the object, giving it a new, dynamic life. The user can now manipulate a voice that is not her own; a special and unfamiliar experience.

It seems that this notion of re-embodiment is qualitatively different from the more "dignified" interactions described within the context of human computer interaction (Spearcons, "Voice-as-Sound"). While in a PC an application is launched and control by voice is only one interaction technique, in objects like Blendie and SonicTexting, the sound behavior is the sole mode of interaction, the voice of the device and its inherent essence.

The remainder of this paper describes the design of an object with a re-embodied voice – one of laughter.

2. THE LAUGHING SWING

The Laughing Swing looks like a simple, regular swing. When you sit on it, it chuckles. As you swing, it laughs, and the higher you go, the harder it chortles. At the peak swinging it is laughing wildly. The swing connects the experience of movement with the experience of laughter. It is a cycle: the user, by swinging, makes the swing laugh, and this laughter causes the user to laugh back

2.1. Motivations

The Laughing Swing, created by Michal Rinott and Michal Rothschild with Leor Weinstein, was conceived as a design exploration into the nature of voice re-embodiment, and specifically of laughter. Laughter is a fascinating phenomenon: on the video website YouTube, a simple video of a baby laughing is one of the 10 most viewed clips of all times, with over 48 million views as of May 2008! [10] We were intrigued by the contagious aspect of laughing, probably familiar to everyone from personal experience. In the Laughing Swing, we sought to create an interactive experience in which the user is induced or even compelled to laugh through this mechanism of contagion. As opposed to the 'static' contagious laughter employed in sitcom "laugh tracks" (i.e. dubbed-in sounds of laughter), our goal was to create a dynamically changing experience in which the user's action influences the laughter behavior.

The Laughing Swing was conceived by considering another familiar aspect of laughter: the feeling of strain in the abdominal muscles when laughing hard, known as the experience of "laughing until my tummy aches" or "being in stitches". This abdominal feeling is somewhat reminiscent of the experience of swinging high, also characterized by a "funny feeling in the stomach" associated with the change in the tightness of the abdominal muscles.

Finally, the Laughing Swing was another exploration, as was SonicTexting, of the nature of interactions which employ tightly coupled motion and sound.

2.2. Components

The Laughing Swing is composed of an accelerometer, a micro-controller, an audio playback component, a memory component, an amplifier, a loudspeaker and a battery pack. All components are housed within the wooden swing seat. The swing acceleration is measured and processed in real time, and a "laughter response curve" is created, generated by dynamically piecing together files with recordings of laughter at different levels.

2.3. Sound design

The swing's laugh was chosen after auditioning a variety of different "laughers". Laughers were people around us known to have funny or contagious laughs. We asked each laugher what makes him or her laugh – and tried to provide such stimulus. Responses ranged from watching a favorite comedy show to being tickled. One strategy we found effective with people who did not know what makes them laugh can be termed "auto-contagion': recording them trying to laugh, and then recording them again while they listen to their own laugh.

The recorded laughter files were analyzed, and the softest (e.g. small chuckle, he he he) and hardest (e.g. hysterical, haaahhh, haahh haah haa ha) bouts of voiced laughter were identified. These were categorized as levels 1 and 6, and all the other laughter bouts were categorized accordingly between levels 2 and 5. The bouts of each level were assembled into a single sound file, with short inter-bout pauses introduced.

Of the eight recorded laughers, the final laughers were selected according to two parameters: the highest laughter level variability, and the availability of sufficient laughter bouts in all six levels. The two selected laughs became the female and male laughter voices of the Laughing Swing.

2.4. Sound Behavior Algorithm

In finding the best way to transition between the different laugh files as a function of acceleration, the simplest approach proved effective. In general, when the (smoothed) acceleration level exceeds a certain value, we move up to the next laughter level. When the acceleration decreases below a certain value (defined with hysteresis), we move down to the previous laughter level. The position of playback in each laughter file is not fixed; rather, we return to the position in which the previous access to this file ended.

It was an initial surprise that this abrupt transition, from one laughter file to a chance position in another file, does not sound like a break or disruption. In fact we found it gives the laughter curve a feeling of authenticity, strengthening the impression that the swing is not laughing on its own but responding to the swinger's behavior. The upward transitions are often a result of the swinger making an intentional push to go higher; in response, the swing abruptly bursts into a strong bout of laughter, or, as the saying goes, "doubles over with laughter". The downward transition, usually not a product of the swinger's explicit action, also sounds natural, like a moment of calming down, an easing down of the laughter intensity. When the transition to a different laughter file accesses the file at a silent inter-bout position, it seems much like a short inhale of air preceding the next laughter bout.
[DEMO1.EXE]

These observations are in line with the scientific research of laughter. In a recent scientific workshop devoted to the phonetics of laughter (in Saarland University, Germany, 2007), the importance of examining the dynamics of laughter not as an isolated sound event, but within interpersonal dialogue, was stressed [11]. Laughter rarely happens with no-one around; in fact, people are about 30 times more likely to laugh when they are in a social situation than when they are alone [12]. In dialog, laughter occurs as a response to another person, and follows the somewhat-unexpected dynamics of a conversation. Therefore it can be anticipated that a laughter curve will sound more natural the more it is attuned – abruptly as may be required - to the interaction with the interlocutor (being, in this case, the swing).

2.5. User Observations

While a thorough user study exceeds the scope of this paper, some informal observations follow.

The Laughing Swing has been exhibited and tried in a range of contexts: a design exhibition (in London, UK), a music festival (in Florence, Italy) and as a as a "sound intervention" within a city center (in Holon, Israel). General responses from a wide range of ages have been of surprise and delight. In the city center, the main audience was children, who were always eager to try the swing. During the three days of this installation, children kept returning and bringing their friends, and there was almost always a queue. We overheard one 5-6 year old child explaining to his mother that there is a small person inside the swing, a fact which did not seem to overly concern him; being, a natural extension of the cartoon world. Generally, children seemed to readily take the laughing behavior into stride.

The swing's physical structure was designed to be intentionally large (80cm width), in order to communicate through its form that it is not a toy for children but rather also intended for use by adults (see Figure 1). In the city center, adults were often hesitant and shy, associating the swing with a childish act, or feeling it to be an overly revealing one to perform in public. However, among those who were tempted to try, a recurring response was that the experience is funny and extremely liberating, bringing back feelings from childhood. The laughter, though "weird", was felt to remove inhibitions.

A different audience altogether used the swing in the music festival in Italy. Here two Laughing Swings (male and female laughs) were installed, and the crowd - composed of teenagers and young adults - used them continuously for the whole two nights of the festival. Some simply enjoyed the experience, while others experimented with the swing and tried to figure out how it worked. A recurring phenomenon were "explainers" - people fascinated with the swing who would stand around it for long periods of time and explain its behavior to passers-by.

Do people laugh when using the Laughing Swing? Some chuckle, some laugh wildly, and some do not. The swing has been displayed to date only in public contexts and only in the midst of (enthusiastic) audiences of people, usually waiting for their turn to swing. Laughing out loud in a situation like this is dependent on many factors: extroversion, mood and affinity to swings being only a few.

On the extroverted side, one behavior that we observed was using the swing as a platform to create a sort of performance for friends and other onlookers. This was done by swinging wildly at the highest possible level, to make the swing hysterical; by mimicking the swing's laughter with their mouths and faces (especially when voice was of the opposite gender); and by hopping off the swing abruptly to try make the laughter stop "in midair".

Figure 1. *The Laughing Swing and a Laughing Swinger*

3. CONCLUSIONS

This author believes that interactions with objects that embody non-verbal voice, and interactions that require the user to produce non-verbal voice, contain promise for new types of dialogue that are lively, intriguing and intimate.

The notion of re-embodiment seems promising for further inquiry. What is the significance of re-embodied sound? Which types of sounds and of objects are 'compatible' for embodiment? When does an object embodied with sound create intimacy? When does it evoke performance?

To further the understanding of this potential, the special characteristics of the human voice, both in phonetic terms and in social communication terms, need to be studied and better understood by those attempting to use non-speech voice as a material for interaction.

4. REFERENCES

[1] D. K. Palladino and B. N. Walker, "Learning Rates For Auditory Menus Enhanced With Spearcons Versus Earcons," in *Proceedings of the 13th International Conference on Auditory Display,* Montréal, Canada, June 26-29, 2007.

[2] T. Igarashi and J. F. Hughes, "Voice as Sound: Using Non-verbal Voice Input for Interactive Control" in *14th Annual Symposium on User Interface Software and Technology, ACM UIST'01,* Orlando, Florida, 2001.

[3] AJ Sporka, SH Kurniawan, P Slavík: Acoustic control of mouse pointer. *Universal Access in the Information Society*, Vol. 4(3):237 – 245, Springer-Verlag, March 2006.

[4] Non-speech Operated Emulation of Keyboard in *Designing Accessible Technology*. London, UK, Springer, 2006, p. 145-154.

[5] T. Holmes, *Electronic and Experimental Music.* Routledge, London, England, 2002.

[6] K. Dobson, B. Whitman and D.P Ellis, "Learning Auditory Models Of Machine Voices" *in Proc of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,* October 16-19, 2005, New Paltz, NY.

[7] G. Pullin, "Hand - Ear Coordination in the Control of Emotionally Expressive Speech Synthesis", Design Report, Interaction Design Lab, Schools of Computing and Design, University of Dundee.

[8] M. Rinott, "SonicTexting", in *Proceedings of the Conference on Human Factors in Computing Systems (CHI '05)* Portland, OR, USA, April 2005, pp. 1144 - 1145 .

[9] M. Hauenstein and T. Jenkins, "Audio Shaker", http://www.nurons.net/audioshaker/about.htm

[10] http://www.youtube.com/browse?t=a&p=1&s=mp&c=0&l=

[11] K. J. Kohler, "Speech-Smile', 'Speech-Laugh', 'Laughter' and Their Sequencing In Dialogic Interaction". In *Proc. of the Interdisciplinary Workshop on The Phonetics of Laughter*, Saarland University, Saarbrücken, Germany August 2007.

[12] R. R. Provine, *American Scientist*, Jan-Feb 1996.